AMENDMENTS TO THE CLAIMS:

1 - 71. (Canceled)

72. (Currently Amended) A method for identifying similar biopolymers ~~biopolymer sequences characterized by a topological pattern of match states~~, the method comprising the steps of:

constructing a statistical model comprising a hidden Markov Model of a set of known sequences that correspond to defined regions of a set of biopolymer sequences to provide ~~characterized by~~ a characteristic topological pattern of match states between the biopolymer sequences, each match state characterized by a scoring matrix, wherein the scoring matrix for a first match state defines a state of similarity for a conserved region of the biopolymer sequences and the scoring matrix for a second match state defines a state of dissimilarity for a divergent region of the biopolymer sequences, [[;]] the model comprising one or more modules of nodes; [[,]]

comparing the set of biopolymer sequences to the statistical model by evaluating ~~the topological pattern of match states of the biopolymer sequences to the topological pattern of match states of the set of known sequences and comparing~~ the scoring matrices of the match states to provide an output score ~~determine the state of similarity or the state of dissimilarity of the biopolymer sequences and the set of known sequences~~; and

~~identifying the biopolymer sequences~~ determining a likelihood that the set of biopolymer sequences is represented by the model and thereby similar biopolymers based on the score ~~state of similarity or the state of dissimilarity with the set of known sequences~~.

73. (Previously Presented) The method of claim 72, wherein the step of constructing the model comprises:

determining the topological pattern of match states of the set of known sequences;

preparing at least one module of nodes for each match state; and

linking the modules of nodes to form the model.

74. (Previously Presented) The method of claim 73, wherein the step of preparing the modules of nodes comprises:

programming the modules of nodes against a training set of data objects characteristic of the topology pattern of match states of the set of known sequences; and

tuning the nodes in an iterative process until the modules encompass the training set of data objects.

75. (Previously Presented) The method of claim 74, wherein the step of programming the modules of nodes comprises defining the scoring matrix for each match state.

76. (Canceled)

77. (Previously Presented) The method of claim 75, wherein the scoring matrix defining a state of dissimilarity is a function of the scoring matrix defining a state of similarity.

78. (Previously Presented) The method of claim 75, wherein the scoring matrix defining a state of dissimilarity is a function of the arithmetic inverse of the scoring matrix defining a state of similarity.

79. (Cancelled)

80. (Previously Presented) The method of claim 72, wherein the set of known sequences consists of two sequences.

81. (Previously Presented) The method of claim 72, wherein the set of known sequences comprises at least three sequences.

82. (Previously Presented) The method of claim 72, wherein the set of known sequences comprises amino acid sequences.

83. (Previously Presented) The method of claim 72, wherein the set of known sequences comprises nucleic acid sequences.

84. (Previously Presented) The method of claim 72, wherein one or more nodes represent an insertion at a first position in the set of known sequences.

85. (Previously Presented) The method of claim 72, wherein one or more nodes represent a deletion at a second position in the set of known sequences.

86. (Previously Presented) The method of claim 72, wherein each node represents a distribution of monomers at defined positions in the set of known sequences.

87. (Previously Presented) The method of claim 86, wherein the distribution of monomers at a first node is different from the distribution of monomers at a second node.

88. (Previously Presented) The method of claim 87, wherein the distribution of monomers is a function of a scoring matrix that relates the distribution of monomers at a first node and a scoring matrix that relates the distribution of monomers at a second node.

89. (Previously Presented) The method of claim 88, wherein the scoring matrix is a function of independent probabilities of a monomer occurrence.

90. (Currently Amended) The method of claim 89, wherein the distribution P(a,b) of monomers a and b, a scoring matrix S(a,b), and independent probabilities of monomers, Q(a) and Q(b) are related such that $S(a,b) = \log(P(a,b) / (Q(a) \cdot Q(b)))$.

91. (Previously Presented) The method of claim 72, wherein the model comprises a first module which characterizes the match state between the set of known sequences in a first region and a second module which characterizes the match state between the set of known sequences in a second region; wherein the match states of the first and second module are different.

92. (Previously Presented) The method of claim 91, wherein the model further comprises a third module that characterizes the match state between the set of known sequences in a third region.

93. (Previously Presented) The method of claim 92, wherein the third module is positioned between the first and second module with respect to the order of the set of known sequences.

94. (Previously Presented) The method of claim 93, wherein the third module indicates similarity between a third region of each set of known sequence, and a sequence profile characterized by altered scoring matrices.

95. (Previously Presented) The method of claim 94, wherein the sequence profile comprises a profile of a modification site.

96. (Previously Presented) The method of claim 95, wherein the modification site is a processing site.

97. (Previously Presented) The method of claim 96, wherein the processing site indicates a preference for at least one basic residue.

98. (Previously Presented) The method of claim 96, wherein the processing site indicates a preference for at least two basic residues.

99. (Previously Presented) The method of claim 96, wherein the processing site comprises a convertase processing site.

100. (Previously Presented) The method of claim 96, wherein the processing site comprises a secretase processing site.

101. (Previously Presented) The method of claim 72, wherein the biopolymer sequences comprise sequences from different species.

102. (Previously Presented) The method of claim 101, wherein the different species comprise mammalian species.

103. (Previously Presented) The method of claim 72, wherein the set of known sequences comprise genomic nucleic acid sequences.

104. (Previously Presented) The method of claim 72, wherein the set of known sequences comprises non-coding regions.

105. (Previously Presented) The method of claim 72, wherein the set of known sequences comprises regulatory regions.

106. (Previously Presented) The method of claim 72, wherein the set of known sequences comprises transcriptional regulatory regions.

107. (Currently Amended) A method for identifying similar biopolymer sequences ~~characterized by a topological pattern of match states~~, the method comprising the steps of:

constructing a statistical model comprising a hidden Markov Model of a set of known sequences that correspond to defined regions of a set of biopolymer sequences to provide ~~characterized by~~ a characteristic topological pattern of match states between the biopolymer sequences, each match state characterized by a scoring matrix, wherein the scoring matrix for a first match state defines a state of similarity for a conserved region of the biopolymer sequences and the scoring matrix for a second match state defines a state of dissimilarity for a divergent region of the biopolymer sequences, the model comprising one or more modules of nodes, each module of nodes representing a different match state, wherein the step of constructing the model comprises programming the modules of nodes against a training set of data objects characteristic of the match state of the set of known sequences, tuning the nodes in an iterative process until the modules encompass the training set of data objects, and linking the modules to form the model;

comparing the set of biopolymer sequences to the statistical model by evaluating ~~the topological pattern of match states of the biopolymer sequences to the topological pattern of match states of the set of known sequences and comparing~~ the scoring matrices for each match state to provide an output score ~~determine the state of similarity or the state of dissimilarity of the biopolymer sequences and the set of known sequences~~; and

~~identifying the biopolymer sequences~~ determining a likelihood that the set of biopolymer sequences is represented by the model and thereby similar biopolymers based on the score ~~state of similarity or the state of dissimilarity with the set of known sequences~~.

108. (Previously Presented) The method of claim 107, wherein the step of programming the modules of nodes further comprises the step of defining a scoring matrix for each match state.

109-110. (Canceled)

111. (Previously Presented) The method of claim 108, wherein the scoring matrix defining a state of dissimilarity is a function of the scoring matrix defining a state of similarity.

112. (Previously Presented) The method of claim 108, wherein the scoring matrix defining the state of dissimilarity is a function of the arithmetic inverse of the scoring matrix defining a state of similarity.

113. (Canceled)

114. (Previously Presented) The method of claim 107, wherein the set of known sequences consists of two sequences.

115. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises at least three sequences.

116. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises amino acid sequences.

117. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises nucleic acid sequences.

118. (Previously Presented) The method of claim 107, wherein one or more nodes represent an insertion at a first position in the set of known sequences.

119. (Previously Presented) The method of claim 107, wherein one or more nodes represent a deletion at defined positions in the set of known sequences.

120. (Previously Presented) The method of claim 107, wherein each node represents a distribution of monomers at defined positions in the set of known sequences.

121. (Previously Presented) The method of claim 120, wherein the distribution of monomers at a first node is different from the distribution of monomers at a second node.

122. (Previously Presented) The method of claim 121, wherein the distribution of monomers is a function of a scoring matrix that relates the distribution of monomers at a first node and a scoring matrix that relates the distribution of monomers at a second node.

123. (Previously Presented) The method of claim 122, wherein the scoring matrix is a function of independent probabilities of a monomer occurrence.

124. (Currently Amended) The method of claim 123, wherein the distribution P(a,b) of monomers a and b, a scoring matrix S(a,b), and independent probabilities of monomers, Q(a) and Q(b) are related such that $S(a,b) = \log(P(a,b) / (Q(a) \cdot Q(b)))$.

125. (Previously Presented) The method of claim 107, wherein the model comprises a first module which characterizes a match state between the set of known sequences in a first region and a second module which characterizes a match state between the set of known sequences in a second region; wherein the match states of the first and second modules are different.

126. (Previously Presented) The method of claim 125, wherein the model further comprises a third module that characterizes a match state between the set of known sequences in a third region.

127. (Previously Presented) The method of claim 126, wherein the third module is positioned between the first and second module with respect to the order of the set of known sequences.

128. (Previously Presented) The method of claim 127, wherein the third module indicates similarity between a third region of each set of known sequence, and a sequence profile characterized by altered scoring matrices.

129. (Previously Presented) The method of claim 128, wherein the sequence profile comprises a profile of a modification site.

130. (Previously Presented) The method of claim 129, wherein the modification site is a processing site.

131. (Previously Presented) The method of claim 130, wherein the processing site indicates a preference for at least one basic residue.

132. (Previously Presented) The method of claim 130, wherein the processing site indicates a preference for at least two basic residues.

133. (Previously Presented) The method of claim 130, wherein the processing site comprises a convertase processing site.

134. (Previously Presented) The method of claim 130, wherein the processing site comprises a secretase processing site.

135. (Previously Presented) The method of claim 107, wherein the biopolymer sequences comprise sequences from different species.

136. (Previously Presented) The method of claim 134, wherein the different species comprise mammalian species.

137. (Previously Presented) The method of claim 107, wherein the set of known sequences comprise genomic nucleic acid sequences.

138. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises non-coding regions.

139. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises regulatory regions.

140. (Previously Presented) The method of claim 107, wherein the set of known sequences comprises transcriptional regulatory regions.

141. (Canceled)

142. (Currently Amended) A computer-readable medium having stored thereon a plurality of instructions, the plurality of instructions including instructions which, when executed by a processor, cause the processor to perform the steps of a method for identifying similar biopolymers ~~biopolymer sequences characterized by a topological pattern of match states~~, comprising ~~of~~:

constructing a statistical model comprising a hidden Markov Model of a set of known sequences that correspond to defined regions of biopolymer sequences in a set of biopolymers to provide ~~characterized by~~ a characteristic topological pattern of match states between the biopolymer sequences, each match state characterized by a scoring matrix, wherein the scoring matrix for a first match state defines a state of similarity for a conserved region of the biopolymer sequences and the scoring matrix for a second match state defines a state of dissimilarity for a divergent region of the biopolymer sequences, [[;]] the model comprising one or more modules of nodes; [[,]]

comparing the set of biopolymer sequences to the statistical model by evaluating ~~the topological pattern of match states of the biopolymer sequences to the topological pattern of match states of the set of known sequences and comparing~~ the scoring matrices of the match states to provide a score ~~determine the state of similarity or the state of dissimilarity of the biopolymer sequences and the set of known sequences~~; and

outputting the score indicative of a likelihood that the set of biopolymer sequences is represented by the model and thereby similar biopolymers ~~identifying the biopolymer sequences based on the state of similarity or the state of dissimilarity with the set of known sequences~~.